AMI: AUGMENTED MULTIPARTY INTERACTION

Steve Renals[†] *and the AMI Consortium*

http://www.amiproject.org

ABSTRACT

AMI is a European Integrated Project (launched in January 2004) that brings together 15 partners from academia and industry. The principal aim of the project is to develop new multimodal technologies that support human interaction in the context of instrumented meeting rooms and remote meeting assistants. The main research themes in AMI include: data collection, annotation and distribution and the development of an infrastructure to support AMI research; meeting dynamics and human-human interaction modelling; multimodal recognition (speech and vision); multimodal integration and content abstraction; and the development of meeting browsers and assistants. Here we present an overview of AMI, with an emphasis on the infrastructure that we are setting up.

1 INTRODUCTION

AMI is concerned with new multimodal technologies to support human interaction, in the context of instrumented meeting rooms and remote meeting assistants. The project aims to enhance the value of multimodal meeting recordings and to make human interaction more effective in real time. These goals will be achieved by developing new tools for computer-supported cooperative work and by designing new ways to search and browse meetings as part of an integrated multimodal group communication, captured from a wide range of devices.

Applications and activities addressed in AMI include multimodal recognition, integration of and coordination among modalities, , meeting dynamics and human-human interaction modelling, the definition of meeting scenarios, content abstraction, and data collection, management, annotation and sharing. In addition to these research activities, AMI has a technology transfer focus, through the exploration and evaluation of advanced end-user applications and prototype systems and a set of training activities (targetted at both students and postdocs) that includes an international exchange programme.

This research will be undertaken in the framework of well-defined and complementary application scenarios: meeting browsers; remote meeting assistants; and integration with wireless presentations. These demonstrators will be developed on the basis of the instrumented meeting room and web-enhanced communication infrastructures, which are already available from some of the AMI partners, and which will be extended in multiple ways.

The project will also make recorded and annotated multimodal meeting data widely available for the research community, thereby contributing to the research infrastructure in the field.

The project began in January 2004. Here we give an overview of the consortium and the research themes on which we will concentrate.

2 CONSORTIUM

AMI is made up of 15 partners (four research institutes, five universities, five companies and a standards representative) and is jointly managed by IDIAP and the University of Edinburgh.

Research Institutes IDIAP; ICSI—the International Computer Science Institute; DFKI; and TNO: Institute of Applied Physics and Human Factors Institute.

Universities University of Edinburgh: CSTR and HCRC; University of Sheffield: SpandH and Information Studies; University of Twente Parlevink group; Munich University of Technology Institute of Man-Machine Communication; Brno University of Technology Institute of Computer Graphics and Multimedia.

Industries Philips Creative Display Systems; Fastcom SA; RealVNC; Spiderphone SA; Novauris.

Standards W3C Interaction domain.

3 RESEARCH THEMES

Depending on the meeting scenario, different information modes will be available, including audio, video, textual, and (possibly) interaction information. To facilitate storing, cross-processing and fusion, and as already initiated in the

[†] Centre for Speech Technology Research, University of Edinburgh, Edinburgh EH8 9LW, UK; s.renals@ed.ac.uk

M4 (MultiModal Meeting Manager) project, all these information sources will be (automatically) synchronised and accurately timestamped. The different information sources available will have to be processed and combined to extract all representation and information that will be required to access the meeting data at different semantic and abstraction levels, through specifically designed database structures and meeting browsers.

The considered audio information will come from headset, lapel and microphone arrays, together with binaural recordings , as well as telephone speech (for web-enhanced communication). Video information will come from recording through multiple cameras , including wide angle and medium-range shots of the participants, and a tabletop 360° view. Further information that we are collecting includes whiteboard activity (including timestamped strokes) and (timestamped) capture of PC presentations capture.

Such an instrumented meeting room has been in use at IDIAP since 2002, and two further meeting rooms (at Edinburgh and TNO) will be making recordings from summer 2004. Based on meeting rooms, the main AMI research themes can be formulated as follows:

1. Incremental development (and distribution across a few AMI partners) of the instrumented meeting room and computer-enhanced conference calling infrastructures.
2. Definition and analysis of possible meeting scenarios, for each of the above applications.
3. Multimodal dialogue modelling and the analysis of human interactions
4. Collection, annotation, and distribution of a large multimodal meetings database recorded in the context of the above scenarios, using the media file server initiated in M4 (see http://mmm.idiap.ch).
5. Speech recognition of overlapped, conversational speech in meetings. Particular challenges include dealing with far-field microphones, and recognizing non-native speakers of English
6. Audio source localization and enhancement: microphone arrays, speaker localization and tracking, segregation of multisource signals
7. Extraction of audio metadata: speaker clustering, speaker turn detection, topic detection, dialogue acts, etc.
8. Analysis and processing of video streams, involving: Face detection and recognition, visual tracking, facial expression recognition, and gesture recognition.
9. Fusion and multimodal recognition, involving: Multimodal integration, audio-visual tracking, multimodal action recognition, multimodal identification of intent and emotion, and multimodal person identification and tracking.
10. Understanding meetings as multimodal interactions, including meeting acts, topic identification, meeting structure modelling and inference, multimodal syntax and multisource decoding, and statistical methods for understanding meetings.
11. Content abstraction, including: structure and segmentation, indexing and retrieval, summarisation, and information extraction.
12. Multimedia presentation: Development of a flexible intelligent information management framework, enabling the structuring of “lower level” streams of multimodal data, and “higher level” metadata obtained from the recognisers applied to the streams. This will thus involve: requirement analysis, browser components, and usability testing.

4 ACKNOWLEDGEMENTS

This work was supported by the European Union FP6 IST Integrated Project AMI (Augmented Multi-party Interaction, IST–FP6–506811, publication AMI-10).